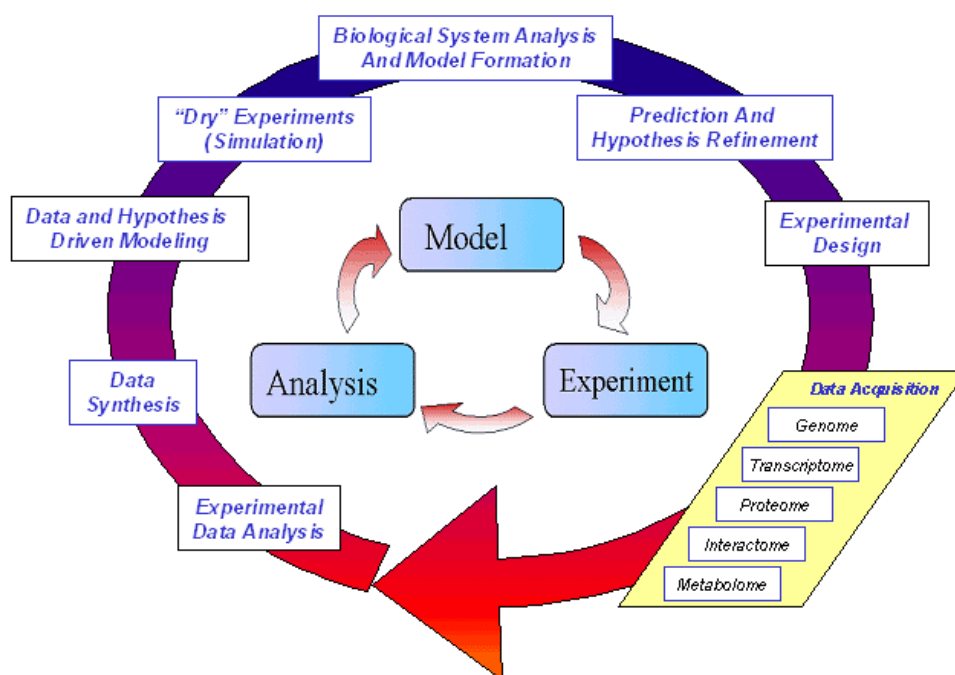




SYSTEMS BIOLOGY IN DRUG DISCOVERY

ΜΑΘΗΜΑ 9^{ου} ΕΞΑΜΗΝΟΥ ΕΜΒΙΟΜΗΧΑΝΙΚΗ ΕΜΒΙΟΙΑΤΡΙΚΗ



ΟΝΟΜΑ ΦΟΙΤΗΤΗ : ΚΑΡΑΓΙΑΝΝΟΠΟΥΛΟΣ ΦΑΕΘΟΝΤΑΣ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ : Λ. ΑΛΕΞΟΠΟΥΛΟΣ

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

Σελ.

1. <u>Εξέλιξη συστημικής βιολογίας</u>	
1.1 Η συστημική βιολογία στην αναπαράσταση του κυττάρου.....	3
1.2 Υπολογιστική μοντελοποίηση στη συστημική βιολογία	4
1.3 Μέθοδοι συλλογής δεδομένων.....	5
1.4 Λάθη σε πειραματικά δεδομένα	8
2. <u>Προσομείωση βιολογικών δικτύων με Γκαουσιανά δίκτυα Μπεις (Bayesian Gaussian Models-Networks)</u>	
2.1 Εισαγωγή στα δίκτυα Bayes.....	9
2.2 Δίκτυα Bayes (Στατικά-δυναμικά).....	10
2.3 Υπολογισμός πιθανότητας εμφάνισης γραφήματος G	12
3. <u>Εφαρμογή γκαουσιανής κατανομής στα δίκτυα Bayes με βάση τα εξαγόμενα αποτελέσματα</u>	
3.1 Μαθηματικό μοντέλο ανάλυσης δεδομένων.....	14
3.2 Αξιολόγηση των αποτελεσμάτων που έχουμε βρει και σύγκρισή τους με τα πραγματικά δεδομένα.....	17
3.3 Το σηματοδοτικό μονοπάτι RAF (RAF signaling pathway).....	19
3.4 Προσομείωση δεδομένων.....	20
4. <u>Ανάπτυξη της συστημικής Βιολογίας με σκοπό την εξατομικευμένη φαρμακευτική θεραπεία</u>	
5. <u>Βιβλιογραφία</u>	25

1. Εξέλιξη συστημικής βιολογίας

1.1 Η συστημική βιολογία στην αναπαράσταση του κυττάρου

Ο όρος «Συστημική Βιολογία» έχει πολύ βαθύ νόημα. Ουσιαστικά, είναι μια προσπάθεια να αναπαρασταθεί η Βιολογία ως σύστημα και να εφαρμόζεται σε αυτή η πλήρης ισχύς των διαφόρων τεχνικών που αναπτύχθηκαν στη φυσική, τη μηχανική και τις φιλοσοφικές επιστήμες. μοντέλα της βιολογίας συστημάτων μπορεί να χρησιμοποιηθούν για την ανακάλυψη φαρμάκων, ώστε να γίνει βελτιστοποίηση των θεραπειών. Μια σειρά από πολύπλοκες ασθένειες οφείλονται στον αποσυντονισμό του συστήματος, και μπορούν να προκληθούν από ένα συνδυασμό περιβαλλοντικών και εσωτερικών παραγόντων. Σε τέτοιες περιπτώσεις, τα κύτταρα θα εξακολουθούν να έχουν φυσιολογικούς φαινοτύπους, ωστόσο η εξασθενημένη σηματοδότηση μεταξύ των κυττάρων μπορεί να είναι επαρκής για να προκαλέσει μια ασθένεια, ακόμη και αν δεν αφήνει σημάδια εντοπισμού αυτής. Η αιτία των ασθενειών αυτών έγκειται στο επίπεδο των συστημάτων. . Στις περιπτώσεις των ασθενειών βιολογίας οι θεραπείες απαιτούν ανάπτυξη στρατηγικών για την κατανόηση, την ταξινόμηση και την μοντελοποίηση των αντίστοιχων μηχανισμών καθώς και της δραστηριότητας των κυττάρων.

Υπάρχουν τουλάχιστον τρία επίπεδα, όπου οι προσεγγίσεις σε επίπεδο συστήματος μπορούν να χρησιμοποιηθούν για την ανάπτυξη φαρμάκων: η λήψη του σήματος,

οι ενδοκυτταρικές αποκρίσεις, και η ενδοκυτταρική επικοινωνία μεταξύ διαφόρων κυτταρικών τύπων. Ένας ουσιαστικός τρόπος αναπαράστασης ενός βιολογικού συστήματος είναι να το διαιρέσουμε σε λειτουργικές ενότητες.

Έχει πρόσφατα αναγνωριστεί ότι ισχυρή μαθηματική αναπαράσταση βιολογικών συστημάτων μπορεί να επιτευχθεί με τη χρησιμοποίηση ενός αριθμός των φιλοσοφικών αρχών (1) . Οποιαδήποτε προσπάθεια να διαμορφωθεί ένα βιολογικό σύστημα αντιμετωπίζει πάντα ένα σημαντικό βαθμό ασάφειας , δεδομένου ότι υπάρχουν πάντα πολλαπλές δυνατότητες για τη μαθηματική αναπαράσταση ενός βιολογικού συστήματος. Η ασάφεια αυτή μπορεί να μειωθεί σημαντικά με την επικέντρωση στον πιο ουσιαστικό πυρήνα βιολογικών διαδικασιών που προσδιορίζονται όταν ένα κύτταρο αναλύεται ως προς τη λειτουργία του και την επιβίωσή του .

Προτείνεται μια καθολική προσέγγιση για την ανάπτυξη ενός βιολογικού συστήματος η οποία συνδέει την ενδοκυτταρική με την εξωκυτταρική δραστηριότητα. Μια τέτοια διαδικασία μπορεί να επιτευχθεί με την εφαρμογή μαθηματικών μοντέλων σε μοριακές μονάδες ενός κυττάρου.

Έχει καταδειχθεί ότι η συμπεριφορά των μεμονωμένων κυττάρων-οργανισμών καθώς και εκείνη των μεμονωμένων κυττάρων των πολυκύτταρων οργανισμών μπορεί να είναι πολύ πολύπλοκη. Επί του παρόντος , ένας μεγάλος αριθμός μεθοδολογιών έχουν αναπτυχθεί για την ανάλυση πολύπλοκων συστημάτων σε διάφορους τομείς της ανθρώπινης δραστηριότητας .

Ένα φιλοσοφικό διάγραμμα για την αναπαράσταση της βιολογίας συστημάτων της κυτταρικής συμπεριφοράς δείχνεται στο ΣΧ. 1 . Το διάγραμμα δείχνει μια σειρά από πιθανές αναπαραστάσεις ενός κυττάρου από έναν αριθμό

γωνιών. Μια ανάπαράσταση, η οποία είναι η πιο εύχρηστη σε επίπεδο βιολογίας, είναι η μορφολογική ανάπαράσταση του κυττάρου, το οποίο δείχνει την εσωτερική πρωτεϊνική σύνθεση, καθώς και άλλα "μέρη". Μια άλλη αποτελεσματική ανάπαράσταση ενός κυττάρου είναι ο συνδυασμός των λειτουργικών μονάδων προς εξασφάλιση της επιβιωσιμότητά τους. Το επόμενο επίπεδο ανάλυσης συνεπάγεται την ύπαρξη διαδικασιών σηματοδότησης μεταξύ των συνδεδεμένων μερών. Το πλήθος των αναπαραστάσεων φαίνεται στο ΣΧ. 1 και σε πολλές περιπτώσεις έχει ήδη αναπτυχθεί σε διάφορους τομείς της μηχανικής, αλλά δεν αναγνωρίζονται ακόμη επαρκώς στη βιολογία.

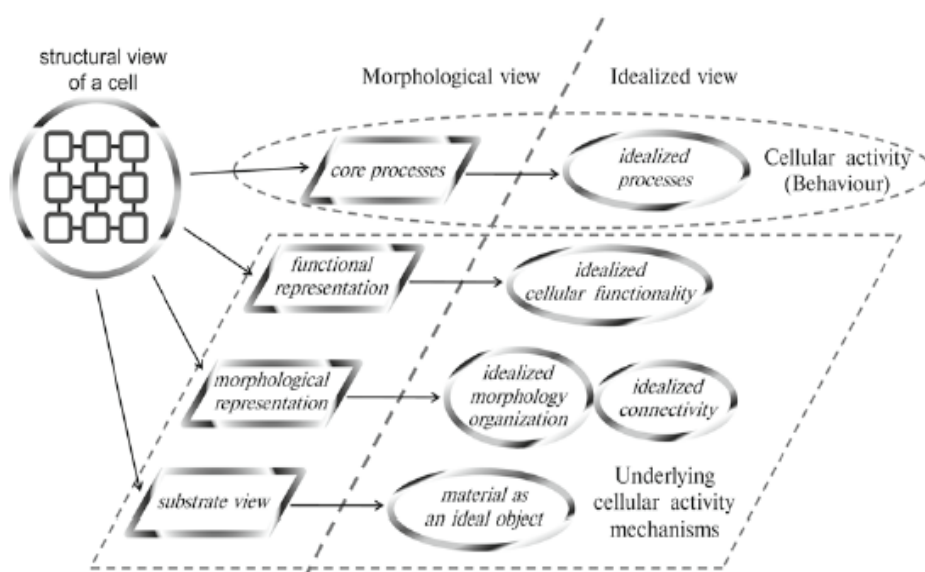


Fig. 1. Functional decomposition of a living cell. A cell is highlighted from a number of angles that can reveal core fundamental processes defining cell behavior. Such decomposition is performed in the morphological and idealized representations. The outlined range of cellular "views" provides guidance for the direction of mathematical model development that would link the molecular mechanism with physiological phenotype.

1.2 Υπολογιστική μοντελοποίηση στη συστημική βιολογία

Οι πρόσφατες πρόοδοι στις πειραματικές τεχνικές της μοριακής βιολογίας έχουν οδηγήσει σε ένα υψηλό ποσοστό παραγωγής ψηφιακών δεδομένων και έχουν τονίσει την ανάγκη για ανεύρεση υπολογιστικών εργαλείων και αλγορίθμων για να βοηθήσουν στην επικύρωση καθώς και τη ανάλυση αυτών. Τα δεδομένα αυτά μπορεί να χρησιμοποιηθούν για την πρόβλεψη νέων δεδομένων και για τη δημιουργία μοντέλων βιολογικών μηχανισμών που εμπλέκουν τη συλλογική λειτουργία χιλιάδων γονιδίων και πρωτεϊνών.

Τέτοιες προσπάθειες κρατούν το κλειδί για την επίτευξη του σκοπού της συστημικής βιολογίας, η οποία έχει ως στόχο να επαναπροσδιορίσει το μέλλον της ιατρικής ενσωματώνοντας τη γνώση από πειραματικές και υπολογιστικές προσεγγίσεις. Αυτά τα υπολογιστικά βήματα είναι απαραίτητα για να καταλάβουμε πώς λειτουργούν τα βιολογικά συστήματα πώς διατηρούν πολλές χαρακτηριστικές ιδιότητες, όπως η αξιοπιστία, και γιατί ορισμένες

διαταραχές ή αλληλουχίες διαταραχών της γενετικής έχουν προεκτάσεις που προκαλούν τη νόσο.

Ενώ οι πειραματικές προσεγγίσεις και ανακαλύψεις έχουν κυριαρχήσει το χώρο της μοριακής βιολογίας κατά το τελευταίο μισό αιώνα, μόνο κατά την τελευταία δεκαετία, λόγω της διαθεσιμότητας τεράστιων βάσεων δεδομένων, έχουν οδηγήσει στην ανάπτυξη μοντέλων για να εξηγηθούν τα ειδικά βιολογικά φαινόμενα. Έτσι δημιουργούνται υποθέσεις, εξάγονται συμπεράσματα, και προτείνονται περισσότερα πειράματα που με τη σειρά τους βοηθούν στη βελτίωση των μοντέλων. Οι ποσοτικές προσεγγίσεις αυτές έχουν συνδράμει στη προγνωστική μοντελοποίηση, που συχνά συνδέεται με τις φυσικές επιστήμες τη μηχανική και τη βιολογία.

Η τεχνολογία των μικροσυστοιχιών (microarray technology) επιτρέπει σε χιλιάδες γονίδια να εκφραστούν σε πολύ υψηλή πυκνότητα, που τα μοτίβα έκφρασης του μελετήθηκαν από την απεικόνιση και την ανάλυση της έντασης της υβριδοποίησης τους. Τα microarrays έχουν ανοίξει ένα νέο τρόπο για να μελετηθεί η έκφραση των γονιδίων σε γονιδιακή κλίμακα. Σε γενετικό επίπεδο, η προσαρμογή στις περιβαλλοντικές διαταραχές γίνεται δυνατή τροποποιώντας τα επίπεδα έκφρασης των διαφόρων γονιδίων. Οι μελέτες των μικροσυστοιχιών περιλαμβάνουν πειραματικό σχεδιασμό, προεπεξεργασία, εξαγωγή, ταξινόμηση και επικύρωση των δεδομένων. Θα πρέπει να σημειωθεί ότι καθένα από αυτά τα πέντε βήματα έχει τη δική του την πολυπλοκότητά και ότι είναι το επίκεντρο της συνεχιζόμενης έρευνας. Για παράδειγμα, ένα από τα μείζονα θέματα με τα ανεπεξέργαστα δεδομένα των μικροσυστοιχιών είναι ότι καταστρέφονται λόγω του θορύβου. Αυτό έχει οδηγήσει σε ενεργό έρευνα στην περιοχή της αφαίρεσης θορύβου από τα δεδομένα και ανάπτυξη μεθόδων για να αναπαρασταθούν οι τιμές που λείπουν.

Υπάρχει ένας αριθμός διαφορετικών τύπων πειραμάτων μικροσυστοιχιών οι οποίες περιλαμβάνουν cDNA με βάση ολιγονουκλεοτίδια και μικροσυστοιχίες πρωτεϊνών. Τα Microarrays χρησιμοποιούνται επίσης ευρέως για την ανακάλυψη φαρμάκων ώστε να δοκιμαστεί η επίδραση των φαρμάκων με σύγκριση της έκφρασης γονιδίων σε φυσιολογικά κύτταρα σε σύγκριση με κύτταρα που έλαβαν θεραπεία με τα αντίστοιχα φάρμακα.

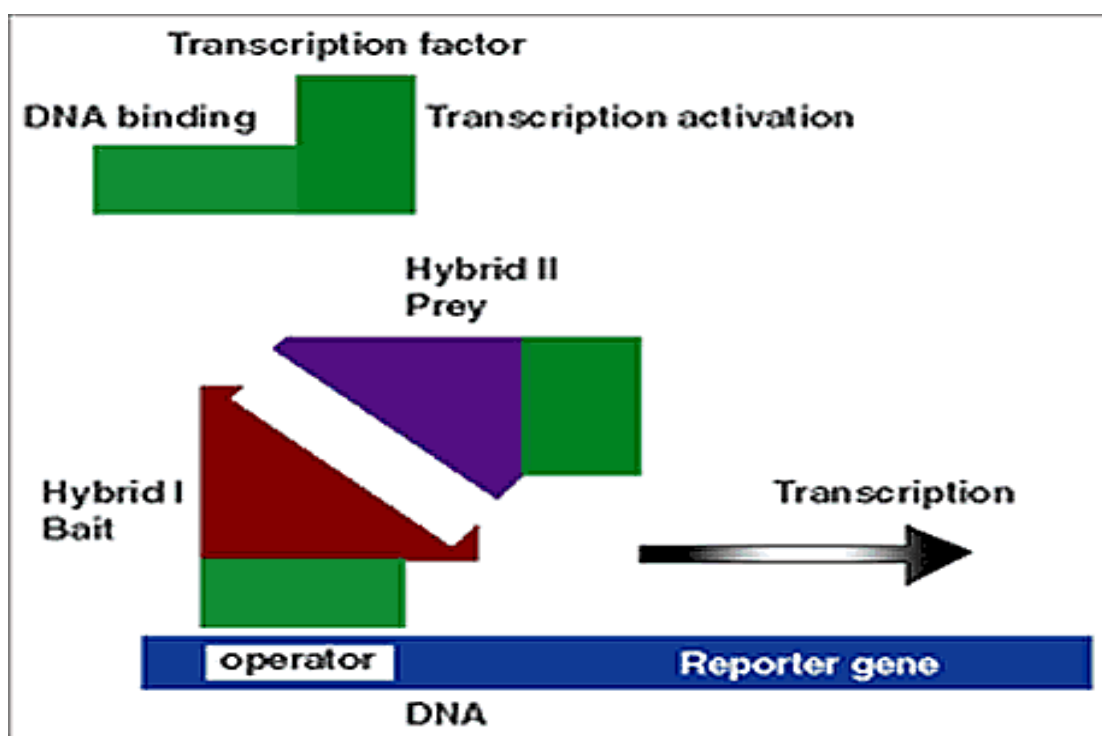
Επίσης είναι ένα ιδιαίτερα σημαντικό εργαλείο για τη μελέτη των θεραπειών για τον καρκίνο, καθώς τα προφίλ έκφρασης του κάθε όγκου τείνουν να είναι διαφορετικά και ως εκ τούτου μπορεί να χρησιμοποιηθεί ως μια μοριακή υπογραφή για την παρακολούθηση της αποτελεσματικότητας της θεραπείας. Η τεχνολογία αυτή έχει επαρκώς εξελιχθεί στην περίπτωση ορισμένων μορφών καρκίνου, επιτρέποντας τα microarrays να χρησιμοποιηθούν για να προβλέψουν το αποτέλεσμα.

1.3 Μέθοδοι συλλογής δεδομένων

Η μεγάλη επιδίωξη της βιολογίας συστημάτων έγκειται στη δυνατότητα μοντελοποίησης της τοπολογίας και της δυναμικής των βιομοριακών δικτύων στην κλίμακα του γονιδίου σε συνδυασμό με την ικανότητα να μετρήσει την συνολική έκφραση των γονιδίων και των πρωτεϊνών. Αυτό μπορεί να

είναι δυνατό μόνο εάν έχουμε πρόσβαση σε αξιόπιστα πειραματικά δεδομένα για τη φύση των δικτύων γονιδιώματος κλίμακας. Όπως είναι γνωστό, ωστόσο, τα περισσότερα πειράματα που αποκαλύπτουν αλληλεπιδράσεις σε αυτή την κλίμακα σε λογικές εργαστηρίου είναι επιρρεπή σε λάθη.

Ας συνοψίσουμε πρώτα τις μεθόδους των πιο κοινών εργαστηριακών πειραμάτων που χρησιμοποιούνται για την παραγωγή του δικτύου δεδομένων. Στη yeast to hybrid method (Y2H) εξετάζουμε δύο γονίδια, το γονίδιο που ρυθμίζει την αντιγραφή (transcription gene) και το γονίδιο το οποίο αντιγράφεται (reporter gene). Το transcription gene αποτελείται από δύο τμήματα, το τμήμα που παράγει την πρωτεΐνη που προσκολλάται (binding domain) στην αλληλουχία ενεργοποίησης του reporter gene (UAS) και το τμήμα (activation domain) που παράγει την πρωτεΐνη που ενεργοποιεί την έκφραση του reporter gene. Όταν τα δύο τμήματα αυτά ενωθούν τότε δίνεται εντολή στο reporter gene να εκφραστεί. Συνεπώς Η ύπαρξη της πρωτεΐνης που παράγεται από το transcription gene υπονοεί την αλληλεπίδραση μεταξύ των πρωτεϊνών Bait-Prey (δόλωμα –κυνηγός)



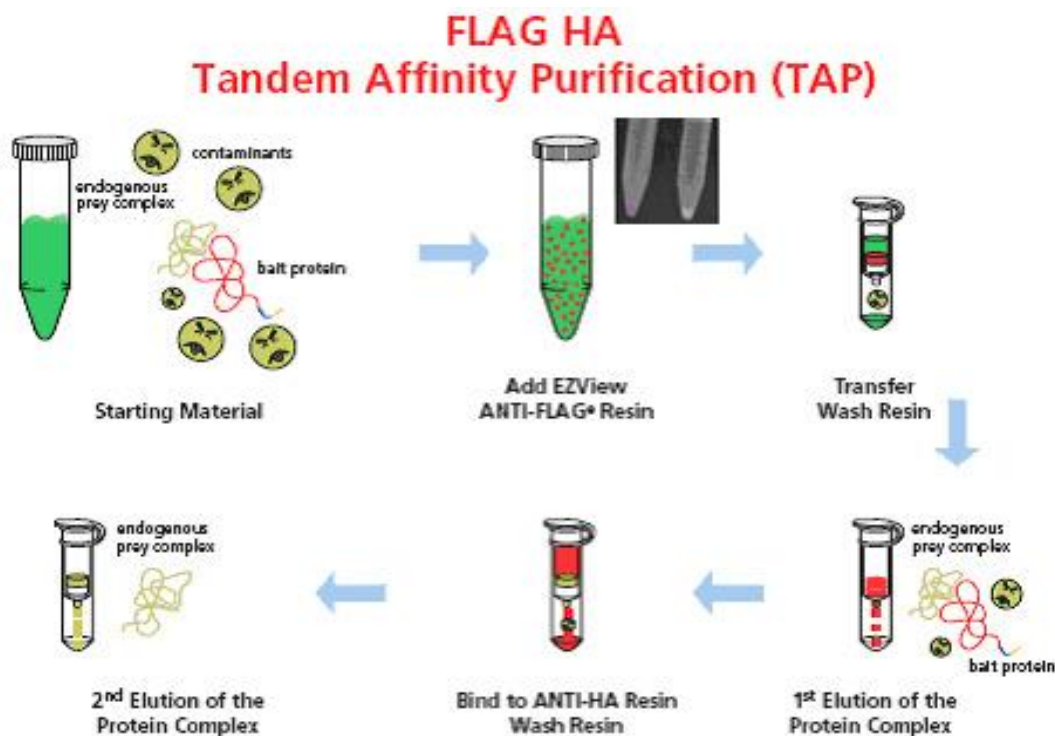
Άλλες μέθοδοι για την ανίχνευση των αλληλεπιδράσεων των πρωτεϊνών είναι:

- AP (Affinity purification) εντοπισμός της πρωτεΐνης από ένα μείγμα με βάση ένα συγκεκριμένο αντίσωμα.
- TAP (Tandem affinity purification) παραλλαγή της μεθόδου AP αλλά με ανίχνευση της πρωτεΐνης σε δύο στάδια καθαρισμού από αντισώματα.

Οι μέθοδοι AP/TAP είναι τεχνικές για τη μελέτη αλληλεπιδράσεων πρωτεΐνης-πρωτεΐνης. Περιλαμβάνουν τη δημιουργία μιας υβριδικής πρωτεΐνης (fusion protein) με την επιθυμητή πρωτεΐνη TAP, στο τέλος. Οι υβριδικές πρωτεΐνες είναι πρωτεΐνες που δημιουργούνται μέσω της ένωσης δύο ή περισσοτέρων γονιδίων τα οποία αρχικά είναι κωδικοποιημένα για ξεχωριστές πρωτεΐνες. Μετάφραση αυτής της πρωτεΐνης υβρίδιο (fusion protein) έχει ως αποτέλεσμα ένα ενιαίο πολυπεπτίδιο με λειτουργικές ιδιότητες που προέρχονται από κάθε μία από τις αρχικές πρωτεΐνες.

Στην αρχική εκδοχή της τεχνικής, η πρωτεΐνη ενδιαφέροντος με την ετικέτα TAP συνδέεται πρώτα σε σφαιρίδια επικαλυμμένα με IgG (αντίσωμα). Εν συνεχεία η πρωτεΐνη TAP διαχωρίζεται με τη βοήθεια ενός ενζύμου, και τελικά ένα διαφορετικό μέρος της πρωτεΐνης TAP δεσμεύεται αναστρέψιμα με σφαιρίδια διαφορετικού τύπου αντισώματος. Αφού η πρωτεΐνη ενδιαφέροντος έχει πλυθεί μέσω δύο στηλών καθαρισμού, μπορεί να εξεταστεί εάν υπάρχουν σε αυτή παράγοντες σύνδεσης. Εάν υπάρχουν παράγοντες σύνδεσης καταλαβαίνουμε πως η πρωτεΐνη ενδιαφέροντος έχει συνδεθεί με τις πρωτεΐνες που έχουμε εισάγει στο μίγμα και πως υπάρχει αλληλεπίδραση μεταξύ αυτών.

Έχουν προταθεί πολλοί συνδυασμοί διαφορετικών πρωτεϊνών ανάλογα με την αλληλεπίδραση που επιθυμούμε να ανακαλύψουμε.



1.4.Λάθη σε πειραματικά δεδομένα

Τα δεδομένα που προκύπτουν περιέχουν λάθη τόσο θετικά (μεγαλύτερη συγκεντρωση πρωτεΐνης) όσο και αρνητικά (μικρότερη συγκεντρωση πρωτεΐνης) σε σύγκριση με την πραγματική πυκνότητα του μείγματος. Η αλληλεπίδραση των πρωτεϊνών in vitro περιέχει απροσδιοριστία και επηρεάζεται από παράγοντες που δεν υφίστανται in vivo. Χαρακτηριστικά:

- Υδρόφιλα υπολείμματα στην επιφάνεια τους αυξάνουν την τάση τους να σχηματίσουν τυχαίες ενώσεις
- Στην μέθοδο TAP ανιχνεύεται η παρουσία μιας πρωτεΐνης σε ένα μόνο πρωτεϊνικό σύμπλεγμα καθόσον χρησιμοποιούμε ένα συγκεκριμένο αντίσωμα
- Επίσης υφίσταται αδυναμία υπολογισμού της περιεκτικότητας πρωτεϊνών που βρίσκονται σε μικρή κλίμακα στο διάλυμα άρα και αρνητικά λάθη

Σε οποιοδήποτε δεδομένο οργανισμό, ο πραγματικός αριθμός των αλληλεπιδράσεων πρωτεϊνών αποτελεί άγνωστη παράμετρος και δε μπορεί να υπολογιστεί παρά μόνο να εκτιμηθεί. Με τα υφιστάμενα υπολογιστικά μοντέλα μπορούμε να υπολογίσουμε τα λάθη που προκύπτουν. Σε κάθε πείραμα κάνουμε μια παρατήρηση και έχουμε τρεις άγνωστες ποσότητες (αριθμός πραγματικών αλληλεπιδράσεων πρωτεϊνών, θετικά λάθη –αλληλεπιδράσεις που έγιναν αλλά δεν εντοπίστηκαν, αρνητικά λάθη-αλληλεπιδράσεις που δεν έγιναν αλλά εντοπίστηκαν).

Ετσι, με 1 HTP πείραμα, έχουμε 3 άγνωστες παραμέτρους (συνολικός αριθμός αληθών αλληλεπιδράσεων και 2 ποσοστά σφάλματος) καθώς και 1 παρατήρηση (αριθμός των παρατηρούμενων αλληλεπιδράσεων). Με 2 πειράματα έχουμε 5 άγνωστες παραμέτρους (συνολικός αριθμός των αλληλεπιδράσεων και 4 ποσοστά σφάλματος) ενώ κάνουμε 3 ανεξάρτητες παρατηρήσεις (αριθμός των αλληλεπιδράσεων που βρέθηκαν από κάθε πείραμα και αριθμός αλληλεπιδράσεων που βρέθηκαν και στα δύο) και με 3 πειράματα έχουμε 7 άγνωστες παραμέτρους και 7 ανεξάρτητες παρατηρήσεις (3 αριθμοί αλληλεπιδράσεων που αναφέρθηκαν από κάθε πείραμα, 3 αλληλεπιδράσεις που αναφέρθηκαν από ακριβώς 2 πειράματα και 1 μέτρηση των αλληλεπιδράσεων που αναφέρθηκε και στα 3 πειράματα). Επιπλέον , έχουμε κάνει την επιπρόσθετη υπόθεση απλούστευσης ότι τα ποσοστά σφάλματος για κάθε πείραμα είναι ανεξάρτητα αλλιώς , υπάρχουν επιπλέον άγνωστες παράμετροι που αντανakλούν εξάρτηση των ποσοστών σφάλματος . Έτσι, σύμφωνα με τις παραδοχές γίνονται, τουλάχιστον 3 ανεξάρτητα πειράματα HTP που απαιτούνται για την εκτίμηση όλων των σχετικών άγνωστων παραμέτρων .

2.Προσομείωση βιολογικών δικτύων με Γκαουσιανά δίκτυα Μπεις (Bayesian Gaussian Models-Networks)

2.1 Εισαγωγή στα δίκτυα Bayes

Στη συστημική βιολογία, υπάρχει αυξημένο ενδιαφέρον για την εκμάθηση των ρυθμιστικών δικτύων και μονοπατιών σηματοδότησης από μεταγονιδιωματικά δεδομένα. Τα δίκτυα Bayes έχουν εφαρμοστεί ευρέως ως ένα δημοφιλές εργαλείο για αυτό

το σκοπό. Η δημοτικότητα τους πηγάζει από την δυνατότητα εύρεσης της παραμετροποιημένης πιθανότητας (βαθμολόγηση) της δομής του δικτύου, η οποία χρησιμοποιείται ευρέως στα δίκτυα Bayes.

Η βαθμολόγηση αυτή βασίζεται σε μια ολοκλήρωση πάνω στο σύνολο του χώρου των παραμέτρων, η οποία είναι μια πολύ δαπανηρή υπολογιστικά διαδικασία. Για να αποκτήσετε την κλειστή μορφή έκφραση της παραμετροποιημένης πιθανότητας, δύο πιθανολογικά μοντέλα έχουν χρησιμοποιηθεί: η πολυωνυμική κατανομή Dirichlet που οδηγεί στη βαθμολόγηση BDE και η γραμμική κατανομή Gauss με αρχική κατανομή Wishart πριν, οδηγώντας στη βαθμολόγηση BGE (4)

Οι προσεγγίσεις αυτές περιορίζονται στο ότι απαιτούν τα δεδομένα διακριτοποιημένα (BDE), ή μπορεί να εντοπίσουν μόνο γραμμικές ρυθμιστικές σχέσεις (BGE). Σε αυτό το κεφάλαιο, θα επικεντρωθούμε στη Gaussian βαθμολόγηση BGE.

Ένα μοντέλο Bayes έχει ως στόχο να μάθει μόνο τις αλληλεπιδράσεις που μπορεί να βρεθούν στα δίκτυα υψηλής βαθμολόγησης, δηλαδή, επιδιώκει να εξαγάγει μόνο τις αλληλεπιδράσεις (συνδέσεις άκμών) που έχουν υψηλή οριακή μεταγενέστερη πιθανότητα.

Η εξαγωγή των ρυθμιστικών δικτύων και των μονοπατιών είναι σημαντική για την ανακάλυψη και ανάπτυξη φαρμάκων, καθώς τα εξαγόμενα μονοπάτια μπορούν να αποκαλύψουν πώς τα γονίδια ή οι πρωτεΐνες ρυθμίζουν το ένα το άλλο. Δηλαδή, μπορεί να φανεί από το εξαγόμενο μονοπάτι πως αλληλεπιδρούν τα γονίδια μεταξύ τους, και έτσι ποια μπορούν να είναι καλοί υποψήφιοι, για τα στοχευμένα φάρμακα

Από την άλλη πλευρά, μπορεί να φανεί ποια γονίδια δεν αλληλεπιδρούν με τα άλλα και δεν έχουν καμία ρυθμιστική επίδραση στα άλλα. Τα γονίδια ή οι πρωτεΐνες

τα οποία δεν έχουν καμία επίδραση σε άλλα είναι λιγότερο πιθανό να είναι «καλοί» υποψήφιοι των φαρμάκων στόχων.

2.2 Δίκτυα Bayes (Στατικά-δυναμικά)

Τα στατικά δίκτυα Bayes (BNs) είναι ευέλικτα μοντέλα για την αναπαράσταση πιθανοτικών σχέσεων μεταξύ των αλληλεπιδρώντων μεταβλητών (π.χ., γονίδια ή πρωτεΐνες). Σε ποιοτικό επίπεδο, το γράφημα ενός BN περιγράφει τις σχέσεις μεταξύ των μεταβλητών X_1, \dots, X_N με τη μορφή υπό όρων σχέσεων εξάρτησης ή ανεξαρτησίας. Σε ποσοτικό επίπεδο, οι τοπικές σχέσεις μεταξύ των μεταβλητών περιγράφονται από δεσμευμένες κατανομές πιθανότητας. Τυπικά, ένα BN ορίζεται από ένα γράφημα G , μια οικογένεια από δεσμευμένες κατανομές πιθανότητας, και τις παραμέτρους τους q , που καθορίζουν από κοινού την κοινή πιθανότητα κατανομής μέσω των μεταβλητών $P(X_1, \dots, X_N | G, q)$.

Το γράφημα G της BN αποτελείται από N κόμβους, που αντιπροσωπεύουν τις μεταβλητές X_1, \dots, X_N , και ένα σύνολο κατευθυνόμενων ακμών που συνδέουν το κόμβους. Το σύνολο των κατευθυνόμενων ακμών υποδεικνύει όρους εξάρτησης μεταξύ των κόμβων. Εάν υπάρχει μια κατευθυνόμενη ακμή από τον κόμβο X_i στον κόμβο X_j , συμβολικά: $X_i \rightarrow X_j$, τότε ο κόμβος X_i λέγεται γονέας (κόμβος) του X_j , και ο X_j ονομάζεται παιδί του κόμβου X_i . Το σύνολο γονέων (κόμβοι) ενός κόμβου X_n , συμβολικά P_n , ορίζεται ως το σύνολο όλων των πατρικών κόμβων του X_n , που είναι το σύνολο όλων των κόμβων από το οποίο κατευθυνόμενες ακμές κατευθύνονται στο X_n εντός του γραφήματος G . Λέμε ότι ένας κόμβος X_n είναι ορφανός αν έχει άδειοπατρικό σύνολο: $p_n = \{ \}$. Αν ένας κόμβος X_k μπορεί να επιτευχθεί ακολουθώντας μια διαδρομή κατευθυνόμενων ακμών ξεκινώντας από τον κόμβο X_i , τότε ο X_k ονομάζεται απόγονος (κόμβος) του X_i και ο X_i λέγεται ένας πρόγονος (κόμβος) του X_k . Η τοπολογία ενός στατικού δικτύου BN ορίζεται να είναι ένα DAG, δηλαδή, ένα κατευθυνόμενο γράφημα στο οποίο κανένας κόμβος δεν μπορεί να είναι δικός του απόγονος ή πρόγονος. Διαγραμματικά, αυτό σημαίνει ότι δεν υπάρχουν κύκλοι κατευθυνόμενων ακμών (loops) στα δίκτυα DAGs.

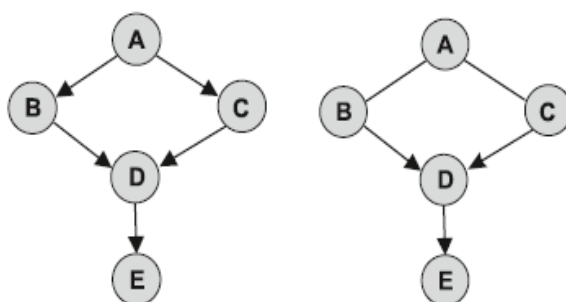


Fig. 1. Example of a Bayesian network with five nodes. *Left panel:* The DAG with five nodes and five directed edges. *Right panel:* The completed partially directed acyclic graph (CPDAG) representation of the DAG shown in the left panel. Two edges of the graph are reversible. These edges are replaced by undirected edges in the CPDAG representation.

Για παράδειγμα, στο DAG που φαίνεται στην αριστερή πλευρά του ΣΧ. 1 , υπάρχουν πέντε μεταβλητές πεδίου συμβολικά A, ..., E. Ο κόμβος A είναι ένας γονέας (κόμβος) και των δύο κόμβων B και C. Οι κόμβοι B και C είναι οι κόμβοι του παιδιού τους A. Δεδομένου ότι υπάρχουν μονοπάτια κατευθυνόμενων ακμών που οδηγούν από τον κόμβο A προς τον κόμβο E, για παράδειγμα: $A \rightarrow B \rightarrow D \rightarrow E$, ο κόμβος A είναι ένας πρόγονος (κόμβος) του E. Το μητρικό σύνολο του κόμβου A είναι το κενό σύνολο, το μητρικό σύνολο του D είναι ίσος $pD = \{B, C\}$, δεδομένου ότι το γράφημα έχει τα δύο άκρα: $B \rightarrow D$ και $C \rightarrow D$ και δεν υπάρχει καμία άλλη ακμή που να δείχνει προς τον κόμβο D. Η κοινή κατανομή πιθανότητας σε στατική δίκτυα Bayes έχει ως εξής:

$$P(X_1, X_2, \dots, X_N / G, q) = \prod_{i=1}^N P(X_i / \pi_i, q_i)$$

όπου $q = (q_1, \dots, q_N)$ είναι ένα διάνυσμα άγνωστων παραμέτρων, και οι γονικοί κόμβοι είναι σύνολα π_i που υπονοούνται από το γράφημα G, συμβολικά: $\pi_i = \pi_i(G)$. Έτσι, κάθε DAG συνεπάγεται μια σειρά από όρους εξάρτησης για τα στατικά δίκτυα BNs. Οι σχέσεις αυτές δίνουν μια μοναδική παραγοντική κατανομή πιθανότητας για τη συνολική δεσμευμένη πιθανότητα. Στην παραγοντοποίηση, κάθε κόμβος X_i εξαρτάται μόνο από τους μητρικούς κόμβους π_i του, και το παραμετρικό διάνυσμα q αποτελείται από N υπο-διανύσματα, έτσι ώστε κάθε υπο-διάνυσμα καθορίζει την τοπική κατανομή πιθανότητας $P(X_i / \pi_i, q_i)$. Για παράδειγμα, η γραφική παράσταση (DAG) που δείχνεται στο αριστερό πλαίσιο του ΣΧ. 1 συνεπάγεται την ακόλουθη πιθανότητα:

$$P(A, \dots, E / G, q) = P(A | q_A) \cdot P(B | A, q_B) \cdot P(C | A, q_C) \cdot P(D | \{A, B\}, q_D) \cdot P(E | D, q_E)$$

Περισσότερα από ένα DAG (directed acyclic graphs-κατευθυνόμενα γραφήματα) μπορεί να συνεπάγονται το ίδιο σύνολο των υπό όρους ανεξαρτησιών και αν δύο DAGs εκχωρούν το ίδιο σύνολο υπό όρους ανεξαρτησιών τότε τα DAGs λέγονται ότι είναι ισοδύναμα. Η σχέση της ισοδυναμίας γραφήματος επιβάλλει ένα σύνολο από κλάσεις ισοδυναμίας πάνω στα DAGs . Τα DAGs μέσα σε μια κλάση ισοδυναμίας έχουν το ίδιο μη κατευθυνόμενο γράφημα, αλλά μπορεί να διαφέρουν ως προς την κατεύθυνση ορισμένων από τις ακμές τους. Δύο DAGs είναι ισοδύναμα αν και μόνο εάν έχουν τον ίδιο σκελετό και το ίδιο σύνολο των ν- δομών. Ο σκελετός ενός DAG ορίζεται ως το μη-κατευθυνόμενο γράφημα το οποίο προκύπτει αγνοώντας όλες τις κατευθυνόμενες ακμές. Μια ν- δομή υποδηλώνει μια διαμόρφωση $X_j \rightarrow X_i \leftarrow X_k$ δύο κατευθυνόμενων ακμών που συγκλίνουν στον ίδιο κόμβο X_i χωρίς ακμή μεταξύ x_j και X_k . Οι κλάσεις ισοδυναμίας των DAGs μπορούν να αντιπροσωπεύονται μοναδικά με τη χρήση (ολοκληρώθηκε), μερικώς κατευθυνόμενων άκυκλων γραφημάτων (CPDAGs) . Μια CPDAG περιέχει τον ίδιο σκελετό όπως η αρχική DAG , αλλά διαθέτει είτε κατευθυνόμενες είτε μη κατευθυνόμενες ακμές. Μια ακμή $X_i \rightarrow X_j$ υπάρχει υποχρεωτικά σε ένα CPDAG όταν όλοι DAGs αυτής της κατηγορίας ισοδυναμίας περιέχουν αυτή κατευθυνόμενη ακμή , ενώ κάθε αναστρέψιμη (μη κατευθυνόμενη) ακμή από το κόμβο X_i στο X_j υποδηλώνει ότι ορισμένοι DAGs στην

κλάση ισοδυναμίας περιέχουν την κατευθυνόμενη ακμή $X_i \rightarrow X_j$, ενώ άλλοι περιέχουν την αντίθετα προσανατολισμένη $X_i \leftarrow X_j$ ακμή. Μια κατευθυνόμενη ακμή σε DAG υποχρεούται στην CPDAG να είναι πάλι κατευθυνόμενη εάν συμμετέχει σε ν-δομή, αλλιώς μπορεί να είναι αναστρέψιμη.

Για παράδειγμα, στην DAG στο αριστερό πλαίσιο του ΣΧ. 1, οι ακμές $B \rightarrow A$ και $C \rightarrow D$ είναι υποχρεωτικά κατευθυνόμενοι, γιατί έτσι και αντιστρέψουμε ένα από

αυτά τα δύο άκρα θα διαγράψουμε τη ν-δομή: $B \rightarrow D, \leftarrow C$. Το άκρο $D \rightarrow E$ είναι επίσης υποχρεωτικά κατευθυνόμενο, καθώς αντιστροφή του θα δώσει δύο νέες ν-δομές, συμβολικά: $B \rightarrow D \leftarrow E$ και $C \rightarrow D, \leftarrow E$. Η CPDAG αναπαράσταση του DAG που φαίνεται στην αριστερή πλευρά του ΣΧ. 1 μπορεί να βρεθεί στο δεξιό πίνακα.

2.3 Υπολογισμός πιθανότητας εμφάνισης γραφήματος G

Λαμβάνοντας υπόψη ένα σύνολο δεδομένων D, η Gaussian βαθμολόγηση BGE μπορεί να χρησιμοποιηθεί για να υπολογίσουμε την εκ των υστέρων πιθανότητα $P(G | D)$ ενός γραφήματος G, δεδομένου του πίνακα αποτελεσμάτων. Για μεταγενέστερη ανάλυση, υποθέτουμε ότι η μήτρα δεδομένων Δέχει μέγεθος $N \times m$ με τις m στήλες να αντιστοιχούν σε

μια ανεξάρτητη υλοποίηση των N μεταβλητών X_1, \dots, X_N . $D_{i,j}$ είναι η j-οστή υλοποίηση του κόμβου X_i και D_p είναι η j-οστή υλοποίηση του μητρικού κόμβου του συνόλου p_i του X_i . Για την εκ των υστέρων πιθανότητα, έχουμε:

$$P(G/D) = \frac{P(G \cap D)}{P(D)} = \frac{P(D/G) * P(G)}{P(D)}$$

Όμως $P(D) = Z_1$ ο πληθάρθρωμος του συνόλου Ω όλων των δυνατών γραφημάτων που σχηματίζονται με τις δεδομένες μεταβλητές. Το Z_1 μεγαλώνει εκθετικά για $n > 6$ και είναι δύσκολο να υπολογιστεί. Αν έχουμε ομοιόμορφη κατανομή των γραφημάτων στο χώρο Ω τότε η πιθανότητα του κάθε γραφήματος είναι:

$$P(G) = \frac{1}{Z_2} * \prod_{i=1}^N \binom{N-1}{|\pi_i|}^{-1}$$

Όπου π_i ο πληθάρθρωμος των πατρικών κόμβων των γραφημάτων G, Z_2 σταθερά εξαρτάται από κατανομή.

Το $P(D/G)$ υπολογίζεται κάνοντας παραμετροποίηση στο χώρο των γραφημάτων G ως προς την παράμετρο q. Έτσι έχουμε:

$$P(D/G) = \int P(D, q/G) dq = \int P(D, q/G) * P(q/G) dq$$

Όμως όπως αναφέραμε παραπάνω:

$$P(D/G, q) = \prod_{i=1}^N \prod_{j=1}^m P(X_i = D_{i,j} / \pi_i = D_{\pi_{i,j}}, q_i)$$

Το παραμετρικό διάνυσμα q αποτελείται από τα διανύσματα q_i $i=1:N$. Επομένως έχουμε:

$$P(q/G) = \prod_{i=1}^N P(q_i/G)$$

Όμως λόγω της υπόθεσης που έχουμε κάνει ότι κάθε παράμετρος εξαρτάται από το συγκεκριμένο κόμβο και ο κόμβος από το πατρικό του σύνολο έχουμε:

$$P(q/G) = \prod_{i=1}^N P(q_i/\pi_i)$$

Έτσι συνολικά έχουμε :

$$P(D/G) = \int \prod_{i=1}^N \prod_{j=1}^m P(X_i = D_{i,j}/\pi_i = D_{\pi_{i,j}}, q_i) * \left(\prod_{i=1}^N P(q_i/\pi_i) dq \right)$$

Και συνολικά η δεσμευμένη πιθανότητα εμφάνισης του γραφήματος G είναι:

$$P(G/D) = \frac{(\prod_{i=1}^N \Psi[D_i^{\pi_i}]) * (\frac{1}{Z_2} * \prod_{i=1}^N \binom{N-1}{|\pi_i|})}{Z_1} = \frac{1}{Z} * \prod_{i=1}^N \Psi[D_i^{\pi_i}] * \binom{N-1}{|\pi_i|}$$

με $\Psi[D_i^{\pi_i}] = \int P(q_i/\pi_i) * (\prod_{j=1}^m P(X_i = D_{i,j}/\pi_i = D_{\pi_{i,j}}, q_i) dq_i$

Το Z που έχει χρησιμοποιηθεί παραπάνω στην εξίσωση είναι ένας παράγοντας κανονικοποίησης

Οι παράγοντες $\Psi[D_i^{\pi_i}]$ εξαρτώνται από το στοχαστικό μοντέλο το οποίο έχουμε χρησιμοποιήσει.

Συνήθως χρησιμοποιείται το γραμμικό μοντέλο Gauss με κανονική κατανομή Wilshart (BGe model) η πολωνυμική κατανομή (Bde model)

ΕΦΑΡΜΟΓΗ ΓΚΑΟΥΣΙΑΝΗΣ ΚΑΤΑΝΟΜΗΣ ΣΤΑ ΔΙΚΤΥΑ BAYES ΜΕ ΒΑΣΗ ΤΑ ΕΞΑΓΟΜΕΝΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

3.1 Μαθηματικό μοντέλο ανάλυσης δεδομένων

Έστω ότι μας δίνεται ένας πίνακας D με τα δεδομένα που προκύπτουν από τις μεταβλητές του πειράματος X_1, X_2, \dots, X_N :

$$D = \begin{pmatrix} D_{1,1} & \cdots & D_{1,m} \\ \vdots & \ddots & \vdots \\ D_{N,1} & \cdots & D_{N,m} \end{pmatrix}$$

Έτσι ώστε $D_{i,j}$ μας δείχνει τα αποτελέσματα (συγκέντρωση) που προκύπτει από την j φορά εκτέλεσης του πειράματος για την μεταβλητή i και η j στήλη $D_{\cdot j} = (D_{1,j}, \dots, D_{N,j})$ μας δείχνει τα αποτελέσματα (συγκεντρώσεις) που προκύπτουν από την j πραγματοποίηση του πειράματος. Στα μοντέλα Gauss για τα δίκτυα Bayes γίνεται η υπόθεση πως τα αποτελέσματα που προκύπτουν από την εκτέλεση του πειράματος

D_j ($j=1, 2, \dots, m$) είναι ένα τυχαίο δείγμα από μια πολυμεταβλητή γκαουσιανή κατανομή $N(\mu, \Sigma)$ με άγνωστο διάνυσμα μέσης τιμής και άγνωστο πίνακα τυπικής απόκλισης Σ . Για την αρχική κοινή κατανομή (joint distribution) του διανύσματος μέσης τιμής μ και του πίνακα ακριβείας $W = \Sigma^{-1}$ γίνεται η υπόθεση πως αποτελούν την κανονική κατανομή Wilshart. Έτσι η δεσμευμένη κατανομή του διανύσματος μ δεδομένου του W είναι $N(\mu_0, (v * W)^{-1})$ με $v > 0$ και η δεσμευμένη κατανομή του W είναι μια κατανομή Wishart με $\alpha > N+1$ βαθμούς ελευθερίας και πίνακα συνδιακύμανσης T_0 , δηλαδή $W(\alpha, T_0)$. Οι Geiger και ο Heckerman απέδειξαν πως η δεσμευμένη πιθανότητα $P(D/G)$ μπορεί να υπολογιστεί με αδύναμες συνθήκες παραμετρικής ανεξαρτησίας σε κλειστή μορφή ως εξής. Ορίζουμε:

$$T_{D,m} = T_0 + S_{D,m} + \frac{v * m}{v + m} * (\mu_0 - \overline{D_m}) * (\mu_0 - \overline{D_m})^T \quad (3)$$

Όπου

$$\overline{D_m} = \frac{1}{m} \sum_{j=1}^m D_j$$

Είναι το μέσο από τα m παρατηρούμενα διανύσματα και

$$S_{D,m} = \sum_{j=1}^m (D_j - \overline{D_m}) * (D_j - \overline{D_m})^T$$

Είναι ο εμπειρικός πίνακας συνδιακύμανσης που προκύπτει από τα υπολογιζόμενα διανύσματα m πολλαπλασιασμένο με $m-1$.

Τα T_0 , μ_0 , α και ν είναι οι υπερπαράμετροι της κανονικής κατανομής Wilshart και πρέπει να καθοριστούν από την αρχή για να γίνει η προσομείωση. Το T_0 είναι ένας πίνακας $N \times N$ το μ_0 είναι ένα $N \times 1$ πίνακας στήλη και τα ν και α είναι βαθμωτοί παράμετροι μίας διάστασης και συνήθως τις ονομάζουμε συνολικούς συντελεστές ακρίβειας.

Φαίνεται από την εξίσωση (3) πως ότι οι υπερπαράμετροι T_0 και μ_0 αναφέρονται στους όρους $S_{D,m}$ και D_m . Έτσι ο πίνακας T_0 και το διάνυσμα στήλη μ_0 αντανakλούν την αρχική πεποίθηση του ερευνητή για τον άγνωστο πίνακα συνδιακύμανσης Σ και το άγνωστο διάνυσμα μ για την κοινή κατανομή των κυρίων μεταβλητών X_1, X_2, \dots, X_N . Εάν η αρχική υπόθεση του ερευνητή είναι πως ο άγνωστος πίνακας συνδιακύμανσης δίνεται από τον αρχικό πίνακα Σ_p τότε μπορεί πχ να θέσει σαν αρχικό πίνακα $\Sigma_p = I_{N \times N}$ όπου ο πίνακας $I_{N \times N}$ είναι ο διαγώνιος πίνακας διαστάσεων N και προτείνεται στην περίπτωση αυτή να χρησιμοποιηθεί σαν αρχικός πίνακας T_0 ο ακόλουθος:

$$T_0 = \frac{\nu * (\alpha - N - 1)}{\nu + 1} * \Sigma_p$$

Με τον ίδιο τρόπο, εάν η αρχική πεποίθηση του ερευνητή είναι πως το άγνωστο διάνυσμα μ μπορεί να δίνεται σε πρώτη φάση από το μ_p , όπου $\mu_p = (0, 0, \dots, 0)^T$, τότε θέτουμε $\mu_0 = \mu_p$. Οι υπερπαράμετροι α και ν αντανakλούν τη βεβαιότητα του ερευνητή για τις αρχικές εκτιμήσεις που έχουμε κάνει για τις υπερπαράμετρους T_0 και μ_0 . Όσο υψηλότερη τιμή έχουν οι συντελεστές α και θ τόσο υψηλότερη είναι η επίδραση των υπερπαράμετρων T_0 και μ_0 στο $T_{D,m}$. Θέτοντας τους αρχικούς παράγοντες ακρίβειας στην μικρότερη τιμή τους δηλαδή θέτοντας $\alpha = N + 2$ και $\nu = 1$ δίνει μια αρχική υπόθεση με μικρή επίδραση στο $T_{D,m}$.

Στα γκαουσιανά μοντέλα (BGe models) η δεσμευμένη πιθανότητα δίνεται όπως αναφέρθηκε ανωτέρω από την ακόλουθη σχέση:

$$P(G/D) = \prod_{i=1}^N \Psi[D_i^{\pi_i}] = \prod_{i=1}^N \frac{P(D^{[X_i, \pi_i]}/G_F(\{X_i, \pi_i\}))}{P(D^{[\pi_i]}/G_F(\{\pi_i\}))} \quad (4)$$

Στον ανωτέρω τύπο το X_i είναι η i μεταβλητή και π_i είναι το σύνολο με τους πατρικούς κόμβους της μεταβλητής X_i στο γράφημα G . $D^{[X_i, \pi_i]}$ και $D^{[\pi_i]}$ είναι οι υποπίνακες του πίνακα δεδομένων D που αποτελούνται μόνο από τις γραμμές που αντιστοιχούν στις μεταβλητές των υποσυνόλων $S_1 = \{X_i, \pi_i\}$ και $S_2 = \{\pi_i\}$. Τα $G_F(\{X_i, \pi_i\})$ και $G_F(\{\pi_i\})$ αντιστοιχούν στα γραφήματα για τα υποσυνόλα των μεταβλητών $S_1 = \{X_i, \pi_i\}$ και $S_2 = \{\pi_i\}$, δηλαδή στα υπογραφήματα με το μέγιστο αριθμό ακμών. Συνολικά γραφήματα δεν επιβάλλουν κάποια ανεξάρτητη συσχέτιση στις μεταβλητές των υποσυνόλων $S_1 = \{X_i, \pi_i\}$ και $S_2 = \{X_i, \pi_i\}$.

Η δεσμευμένη πιθανότητα για ένα υποσύνολο δεδομένων $D^S \subseteq D$, το οποίο αποτελείται από m εκτελέσεις του N διαστάσεως υποσυνόλου $S \subseteq \{X_1, X_2, \dots, X_N\}$ μπορεί να υπολογιστεί όταν το συνολικό γράφημα G δίνεται, από τον ακόλουθο τύπο:

$$P(D^S/G_F(S)) = (2\pi)^{-N*\frac{m}{2}} * \left\{ \frac{v}{v+m} \right\}^{\frac{N}{2}} * \frac{c(N,a)}{c(N,a+m)} * \det(T_0^S)^{\frac{a}{2}} * \det(T_{D,m}^S)^{-(a+m)/2}$$

Όπου $\det(T_0^S)$ και $\det(T_{D,m}^S)$ είναι οι ορίζουσες των υποπινάκων T_0^S και $T_{D,m}^S$ και αποτελούνται μόνο από τις γραμμές και τις στήλες που αντιστοιχούν στις μεταβλητές του υποσυνόλου S . Η ποσότητα $T_{D,m}$ έχει αναφερθεί ανωτέρω πως υπολογίζεται και σχετικά με τις ποσότητες $c(N,a)$ και $c(N,a+m)$ έχουμε:

$$c(N,a) = \{2^{a*\frac{N}{2}} * \pi^{N*\frac{N-1}{4}} * \prod_{l=1}^N \Gamma(\frac{a+1-l}{2})\}^{-1}$$

Αν και τα περισσότερα διαθέσιμα βιολογικά δίκτυα και τα μονοπάτια δεδομένων είναι παθητικά παρατηρήσιμα (observational data) μερικές φορές οι ερευνητές μπορούν να επέμβουν εξωτερικά και να μεταβάλλουν κάποιες παραμέτρους χρησιμοποιώντας για παράδειγμα ορισμένες ουσίες που απενεργοποιούν ή ενεργοποιούν την έκφραση ορισμένων γονιδίων. Τα δεδομένα αυτά είναι αυτά που προκύπτουν με παρεμβολή και ονομάζονται interventional data. Εάν οι παρεμβολές αυτές είναι ιδανικές τότε οι τιμές των ποσοτήτων που προκύπτουν για το σύστημα δεν εξαρτώνται από τις εσωτερικές παραμέτρους του συστήματος αλλά μόνο από τις εξωτερικές συνθήκες τις οποίες επιβάλλουμε. Ωστόσο η συγκέντρωση ορισμένων μεταβλητών τις οποίες ελέγχουμε εξωτερικά μπορεί να επηρεάσει τη συγκέντρωση των άλλων μεταβλητών. Έτσι με τον τρόπο αυτό μπορούμε να διαπιστώσουμε την αιτιώδη σχέση μεταξύ των μεταβλητών (directed edges). Κάτω από ικανοποιητικές συνθήκες ένας συνδυασμός από παρατηρούμενα δεδομένα και δεδομένα με παρεμβολή μπορεί να αναλυθούν χρησιμοποιώντας δίκτυα Bayes.

Οι τοπικές τιμές των παραγόντων $\Psi[D_i^{\pi_i}]$ στη δεσμευμένη πιθανότητα της σχέσης (4) προκύπτουν για κάθε μεταβλητή X_i από τα δεδομένα στα οποία δεν είχαμε παρεμβολή. Έτσι ορίζουμε το διάνυσμα στήλη διαστάσεων $m \times 1$, I το οποίο περιέχει τις τιμές $I(j)$ με τον ακόλουθο τρόπο: $I(j)=i$ όταν στη μεταβλητή X_i έγινε παρεμβολή κατά τη j παρατήρηση και έτσι γίνεται αντικατάσταση των δεδομένων στη σχέση (4) με τον ακόλουθο τύπο:

$$P(G/D) = \prod_{l=1}^N \Psi[D_{i,l}^{\pi_i}]$$

Όπου $D_{i,l}^{\pi_i} = \{D_{i,j}, D_{\pi_i,j} / j \in \{1,2,\dots,m\} : I(j) \neq i\}$ και είναι το υποσύνολο των δεδομένων D με συνολικό πλήθος $m_i \leq m$ για τις m_i φορές που έγινε η μέτρηση στη μεταβλητή X_i και το σύνολο των πατρικών του κόμβων π_i τότε και μόνο τότε όταν δεν είχαμε παρεμβολή στο κόμβο I με $I(j) \neq i$.

Οι συγκεντρώσεις οι οποίες θα χρησιμοποιήσουμε για να υπολογίσουμε τις τοπικές τιμές των παραγόντων $\Psi[D_i^{\pi_i}]$ διαφέρουν από κόμβο σε κόμβο. Για παράδειγμα για τον κόμβο X_i πρέπει να αντικαταστήσουμε τον αρχικό πίνακα D από τον υποπίνακα $D(i)$ ο

οποίος έχει διαστάσεις $N \times m$. Ο πίνακας $D(i)$ μπορεί να εξαχθεί από τον πίνακα D αφαιρώντας όλες τις στήλες στις οποίες όπου έχει γίνει παρεμβολή στη μεταβλητή X_i . Έτσι πρέπει για $i=1 \dots n$ να αφαιρέσουμε όλες τις στήλες $j \in \{1, 2, \dots, m\}$ με $I(j)=I$ από τον αρχικό πίνακα D και έτσι προκύπτει ο πίνακας $D(i)$. Εν συνεχεία πρέπει να κατασκευαστούν πίνακες $T_{D(i)}$, ένας για κάθε μεταβλητή X_i οι οποίες θα υπολογιστούν από τους πίνακες $D(i)$. Έτσι με ανάλογο τρόπο όπως προέκυψαν οι σχέσεις για τις μετρήσεις από παρατήρηση (observational data) έχουμε:

$$T_{D(i),m} = T_0 + S_{D(i),m} + \frac{v * m_i}{v + m_i} * (\mu_0 - \bar{D}(i)_m)^- * (\mu_0 - \bar{D}(i)_m)^T$$

$$P(G/D) = \prod_{l=1}^N \Psi[D_{i,l}^{\pi_l}] = \prod_{l=1}^N \frac{P(D^{[X_i, \pi_l]} / G_F(\{X_i, \pi_l\}))}{P(D^{[\pi_l]} / G_F(\{\pi_l\}))} \quad (4)$$

Και για τις m φορές εκτέλεσης της μέτρησης για τον πληθάρημο N^0 του συνόλου $S \subseteq \{X_1, X_2 \dots X_N\}$ έχω :

$$P(D(i)^S / G_F(S)) = (2\pi)^{-N * m_i / 2} * \left\{ \frac{v}{v + m_i} \right\}^{\frac{N}{2}} * \frac{c(N^0, a)}{c(N^0, a + m_i)} * \det(T_0^S)^{\frac{a}{2}} * \det(T_{D(i),m_i}^S)^{-(a+m_i)/2}$$

Όπου $G_F(S)$ είναι ένα πλήρες γράφημα για το υποσύνολο S των μεταβλητών με πληθάρημο N^0 και $T_{D(i),m}^S, T_0^S$ τα οποία αποτελούνται μόνο από N^0 γραμμές και στήλες που αντιστοιχούν στις μεταβλητές του υποσυνόλου S .

3.2 Αξιολόγηση των αποτελεσμάτων που έχουμε βρει και σύγκρισή τους με τα πραγματικά δεδομένα

Έστω $G = \{G_1, G_2 \dots G_T\}$ το σύνολο των πιθανών γράφων που παράγονται από τις μεταβλητές X_1, X_2, \dots, X_N . Για να αξιολογήσουμε τη μεταγενέστερη δεσμευμένη πιθανότητα μιας ακμής F η οποία υπάρχει στο διάγραμμα που έχουμε υπολογίσει ότι έχει τη μεγαλύτερη πιθανότητα να είναι το σωστό, ορίζουμε τον ακόλουθο συντελεστή :

$$\widetilde{P}(F/D) = \frac{1}{T} \sum_{i=1}^T I_F(G_T)$$

Ο συντελεστής αυτός είναι ένας δυαδικός συντελεστής και παίρνει την τιμή 0 όταν η ακμή του τελικού διαγράμματος δεν υφίσταται στο γράφημα G_T ενώ παίρνει την τιμή 1 όταν η ακμή αυτή υπάρχει στο γράφημα G_T .

Όταν το τελικό γράφημα είναι γνωστό τότε χρησιμοποιούνται οι καμπύλες ROC (Receiver operator characteristic curves) προκειμένου να αξιολογηθεί το ποσοστό ανακατασκευής του δικτύου. Χρησιμοποιούμε το δείκτη $e_{i,j} = 1$ όταν στο πραγματικό δίκτυο υπάρχει συσχέτιση μεταξύ των κόμβων i και j είτε έχουμε βελάκι με κατεύθυνση είτε όχι (ανάλογα πως θέλουμε να ελέγξουμε το δίκτυο) και $e_{i,j} = 0$ αν δεν υπάρχει συσχέτιση και εν συνεχεία σχηματίζουμε το δείκτη $\varepsilon(\theta)$ με:

$$\varepsilon(\theta) = \{ e_{i,j} \mid \widetilde{P}(F/D) > \theta \}$$

το οποίο μας δείχνει όλες τις ακμές στο πιο πιθανό γράφημα που προκύπτει πως ξεπερνούν ένα συγκεκριμένο όριο θ . Θεωρώντας ένα δεδομένο όριο θ μπορούμε εν συνεχεία να υπολογίσουμε τον αριθμό των αληθών θετικών (true positive -TP), λανθασμένων θετικών (false positive - FP) και των λανθασμένων αρνητικών (false negative-FN) συχετίσεων και εν συνεχεία να υπολογίσουμε την ευαισθησία της μέτρησης

$$\text{sensitivity: } S = \frac{TP}{TP+FN}$$

$$\text{inverse specificity: } I = \frac{FP}{TN+FP}$$

Όμως αντί να επιλέξουμε μια συγκεκριμένη τιμή του θ παίρνουμε ένα τυχαίο διάστημα του θ ως όριο και ολοκληρώνουμε και τελικά σχεδιάζουμε τη sensitivity με την inverse specificity σε δύο άξονες. Έτσι προκύπτει η καμπύλη ROC. Ένα ποσοτικό μέγεθος για την καμπύλη ROC έτσι ώστε να γίνει εκτίμηση της ακρίβειας αυτής μπορεί να προκύψει ολοκληρώνοντας και παίρνοντας το εμβαδόν του όγκου που περικλείει. Αυτό είναι γνωστό σαν τιμή AUC και είναι ένα μέτρο της πιστότητας της αναπαράστασης του γραφήματος. Τιμή 1 είναι η μέγιστη και είναι το πάνω όριο όταν υπολογίζουμε επακριβώς το γράφημα G ενώ η ελάχιστη τιμή της είναι 0,5 όπου το γράφημα είναι τελείως τυχαίο.

Ένα ακόμα πιο άμεσο και πιο κατανοητό κριτήριο είναι δίνεται από το λόγο αληθών θετικών συχετίσεων προς λανθασμένων θετικών ($TP/FP=5$). Αυτό το κριτήριο μας δείχνει ποια είναι η αναλογία μεταξύ αληθών θετικών συχετίσεων προς λανθασμένων θετικών και συνήθως παίρνουμε σαν κάτω αποδεκτό όριο το 5.

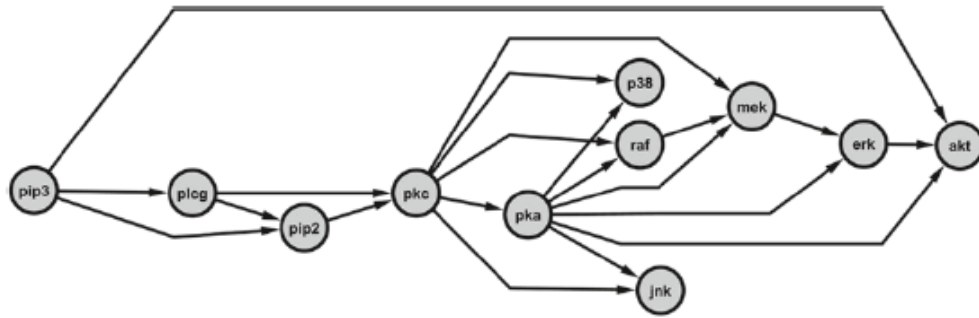


Fig. 3. The RAF-pathway as presented by Sachs et al. (15).

where $G_F(S)$ is a full graph for the subset S of variables of cardinality N^* , and T_0^S and $T_{D(i),m-1}^S$ are sub-matrices of T_0^S and $T_{D(i),m}^S$ consisting only of those N^* rows and columns that correspond to variables in the subset S .

3.3 Το σηματοδοτικό μονοπάτι RAF (RAF signaling pathway)

Το σηματοδοτικό μονοπάτι RAF το οποίο φαίνεται στο ανωτέρω σχήμα είναι ένα βιολογικό σύστημα το οποίο έχει μελετηθεί διεξοδικά και περιγράφει τις διακυτταρικές συσχετίσεις ανάμεσα στα διαφορετικά μόρια που αλληλεπιδρούν για τη μετάδοση του σήματος. Κατά τη διάρκεια μεταφοράς του σήματος στις κυτταρικές πρωτείνες ειδικά ένζυμα (κινάση) τροποποιούν συγκεκριμένες πρωτείνες στόχους προσθέτοντας σε αυτές φωσφορικές ομάδες. (φωσφορισμός). Αυτό οδηγεί σε λειτουργικές αλλαγές των στόχων έτσι ώστε εν συνεχεία πραγματοποιούνται περαιτέρω χημικές αντιδράσεις στην αλυσίδα μεταφοράς της πληροφορίας. Καθώς οι κινάσες έχει ανακαλυφθεί πως ρυθμίζουν τα κυτταρικά μονοπάτια η απορύθμιση αυτών μπορεί να προκαλέσει ασθένειες ακόμα και καρκίνο. Για το λόγο αυτό μελετήθηκαν τα επίπεδα έκφρασης 11 φωσφορούχων πρωτεϊνών και φωσφολιπιδίων σε 1200 άτομα με βάση την κυτομετρία (cytometry).

Πέραν των στοιχείων που πήραμε με παρατήρηση μετρήθηκαν και τα επίπεδα των 11 μορίων μετά από 9 διαφορετικές διεγέρσεις που έγιναν από άλλα κύτταρα. Συγκεκριμένα οι μετρήσεις των ουσιών στο μίγμα έγιναν 15 λεπτά αφότου έγινε η διέγερση αυτών με τα 9 μόρια. Από τα 9 μόρια που χρησιμοποιήσαμε τα 3 μόρια δεν έχει αποδειχθεί πως προκαλούν κάποια αλλαγή στις συγκεντρώσεις των 11 μορίων. Επίσης τα υπόλοιπα 6 μόρια μπορούν να θεωρηθούν ιδανικά δηλαδή επηρεάζουν μονάχα ένα μόριο από το μονοπάτι RAF όπως έχει μελετηθεί στο παρελθόν και ξέρουμε πολύ αναλυτικά τις επιδράσεις αυτές όπως φαίνονται στον παρακάτω πίνακα

Από τα 1200 πειραματικά δεδομένα που συλλέξαμε από τους ασθενείς χρησιμοποιήσαμε μόνο τις 500 τις οποίες τις οργανώσαμε σε 5 ομάδες των 100 δειγμάτων. Επίσης τα δεδομένα που προκύπτουν από την εισαγωγή των εξωτερικών παραγόντων τα οργανώνουμε σε μία ομάδα των 100, $m=100$ η οποία αποτελείται από 14 μετρήσεις των 6 επιδράσεων και 14 μετρήσεις από τα δείγματα που πήραμε από τους ανθρώπους. Δεδομένου πως το δίκτυο RAF είναι σαφώς καθορισμένο θα χρησιμοποιήσουμε τα ποσοστά που προκύπτουν για να επιβεβαιώσουμε την ορθότητα των δικτύων Bayes.

3.4 Προσομείωση δεδομένων

Στις προσομοιώσεις που κάναμε χρησιμοποιήσαμε το Bayes Gauss model το οποίο αναλύθηκε ανωτέρω. Αυτό έγινε σε δύο στάδια στο πρώτο στάδιο αναλύθηκαν τα αποτελέσματα από τα observational data (απλή παρατήρηση) ενώ στο δεύτερο τα δεδομένα από τα interventional data (αποτελέσματα με παρεμβολή) τα δεδομένα αρχικά θεωρήθηκε πως υπακούουν την κατανομή Wilshart και χρησιμοποιήσαμε σα μεταβλητές στα πειράματά μας όπως είχαμε αναφέρει παραπάνω : $v=1$, $\alpha=N+2$, $\mu_0=(0,\dots,0)^T$ και $I_0=0.5*I_{N \times N}$. Σε ορισμένες παρεμβολές που κάναμε στα μόρια με τις 6 ουσίες τα αποτελέσματα που πήραμε δεν ήταν σύμφωνα με τα προβλεπόμενα. Δηλαδή σε ορισμένα δείγματα τοποθετήσαμε αναστολές δράσης των ουσιών όμως στα προϊόντα υπήρχε μια ποσότητα της ουσίας. Επίσης σε ορισμένα άλλα πειράματα εισάγαμε ενεργοποιητή στο δείγμα αλλά αυτό δεν οδήγησε σε υψηλές περιεκτικότητες της ουσίας που ενεργοποιείται. Για να το αντιμετωπίσουμε αυτό στα αποτελέσματα που πήραμε θεωρήσαμε μόνο τα δείγματα που είχαν την μικρότερη η τη μεγαλύτερη αντίστοιχα συγκέντρωση. Έτσι προέκυψαν τα ακόλουθα αποτελέσματα:

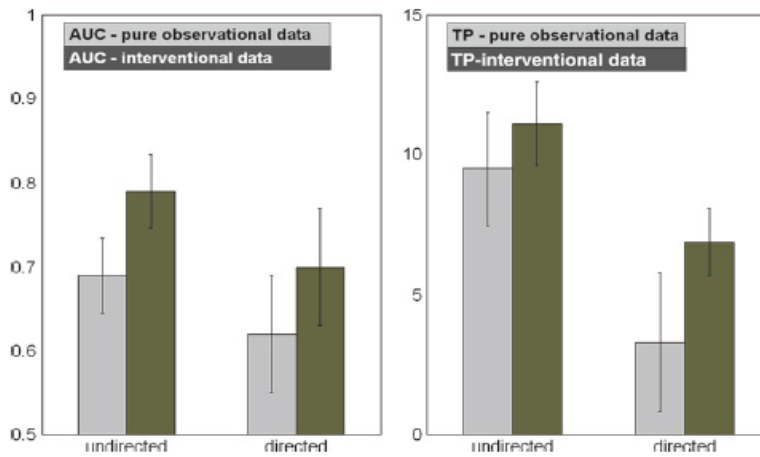


Fig. 4. AUC and TP count histograms. Summary of average AUC (left panel) and (TP/FP = 5) values (right panel) for the observational and interventional flow cytometry data from the RAF-pathway. The network shown Fig. 3 was used to evaluate the correctness of the extracted *undirected* and *directed* edge relation features. The bars in the histograms represent average AUC values and (TP/FP = 5) counts across five independent data sets each. The error bars of the histograms correspond to one standard deviation.

Δεδομένου πως το αρχικό δίκτυο που θέλουμε να προσεγγίσουμε είναι γνωστό τα αποτελέσματα μπορούμε εύκολα να τα διασταυρώσουμε για την ορθότητα της ακρίβειας υπολογίζοντας τους συντελεστές AUC και TF/FP. Στα δίκτυα τα οποία έχουμε υπολογίζουμε δύο συντελεστές ο πρώτος για τα δίκτυα που προκύπτουν με κατευθυνόμενα άκρα (DAG) όσο και τα δίκτυα που προκύπτουν με μη κατευθυνόμενα άκρα (CPDAG).

Η γενική αίσθηση που προκύπτει από τα ιστογράμματα που φαίνονται στο ανωτέρω σχήμα είναι πως τα δίκτυα με παρεμβολή (έγχρωμα δίκτυα) εμφανίζουν καλύτερα ποσοστά ανακατασκευής του δικτύου σε σύγκριση με τα δίκτυα με απλή παρατήρηση (ασπρόμαυρα δίκτυα). Επιπρόσθετα από τα ιστογράμματα μπορούμε να εκτιμήσουμε πως η ανακατασκευή των δικτύων με μη κατευθυνόμενα άκρα δίνει καλύτερους συντελεστές AUC και TF/FP σε σύγκριση με τα δίκτυα με κατευθυνόμενα άκρα. Δηλαδή από τα δεδομένα με παρεμβολή υπάρχει πολύ μεγαλύτερη πληροφορία σε σύγκριση με τα δεδομένα με απλή παρατήρηση. Επίσης το να ανακατασκευάσουμε

το δίκτυο με μη κατευθυνόμενα άκρα είναι πιο εύκολο σε σύγκριση με το να το ανακατασκευάσουμε με κατευθυνόμενα άκρα.

Η υψηλότερη ακρίβεια που λαμβάνουμε από τα δεδομένα μας είναι AUC 0,8 και επιτυγχάνεται όταν προσεγγίζουμε το δίκτυο με μη κατευθυνόμενες ακμές από τα δεδομένα με εξωτερική παρεμβολή. Ο συντελεστής αυτός για τα δεδομένα με απλή παρατήρηση είναι μικρότερος από 0,7 που είναι ακόμα μικρότερος από το AUC για την εύρεση των κατευθυνόμενων άκρων από τα δίκτυα με παρεμβολή (AUC=0,71). Επίσης ο μικρότερος συντελεστής AUC ,0,63, προκύπτει για τις κατευθυνόμενες ακμές από τα απλά δεδομένα με παρατήρηση.

Ενώ οι τιμές του AUC είναι δύσκολο να ερμηνευτούν και να γίνουν κατανοητές από το χρήστη πηγαίνουμε στον άλλο δείκτη ($TP/FP = 5$) δηλαδή πόσες αληθείς ακμές TP παίρνουμε όταν έχουμε ψευδείς ακμές 5. Οι μη κατευθυνόμενες ακμές αυξάνουν μερικώς από 9,6 στα απλά δεδομένα σε 11,2 στα δεδομένα με παρεμβολή. Αντίθετα ο αριθμός των κατευθυνόμενων ακμών αυξάνεται από 3,7 στα απλά δεδομένα σε 7,1 στα δεδομένα με παρεμβολή. Η αύξηση αυτή είναι σημαντική και μεγάλη γιατί βλέπουμε πως ο αριθμός ουσιαστικά διπλασιάζεται. Ωστόσο ακόμα και 7,1 σωστά κατευθυνόμενα άκρα δεν είναι κάποιος εντυπωσιακός αριθμός καθώς για αυτά προκύπτουν 5 λάθος κατευθυνόμενες ακμές.

Δεδομένου ότι υπάρχουν 20 αληθώς κατευθυνόμενες ακμές στο μονοπάτι RAF- που φαίνεται στο Σχ. 3, 7.1 αληθώς κατευθυνόμενες ακμές αντιστοιχούν σε ευαισθησία μόνο 35 τοις εκατό. Από την άλλη πλευρά, υπάρχουν 90 ψευδώς θετικά μη κατευθυνόμενες ακμές στο μονοπάτι RAF ώστε οι πέντε λανθασμένες θετικές (FP) αντιστοιχούν σε μια specificity σχεδόν 95 τοις εκατό. Ωστόσο, το παράδειγμα αυτό αποδεικνύει ότι η Μπαεσιανή μεθοδολογία μπορεί να βοηθήσει στην αποσαφήνιση των σχέσεων μεταξύ των ρυθμιστικών πρωτεϊνών με χρήση κυτανομετρίας ροής δεδομένων. Σημειώνουμε ότι έχει γίνει η προσομείωση για μεγαλύτερο δείγμα $m = 5400$ σημεία επεμβατικών δεδομένων και ότι λαμβάνεται μια πολύ μεγαλύτερη ακρίβεια ανασυγκρότησης του δικτύου.

4. ΑΝΑΠΤΥΞΗ ΤΗΣ ΣΥΣΤΗΜΙΚΗΣ ΒΙΟΛΟΓΙΑΣ ΜΕ ΣΚΟΠΟ ΤΗΝ ΕΞΑΤΟΜΙΚΕΥΜΕΝΗ ΦΑΡΜΑΚΕΥΤΙΚΗ ΘΕΡΑΠΕΙΑ

Η συστημική βιολογία και η φαρμακογονιδιωματική (pharmacogenomics) αναδύονται και υπόσχονται σε πολλά πεδία την παροχή μια σε βάθος κατανόησης των ασθενειών που θα επιτρέψουν την εξατομικευμένη θεραπεία. Μελέτες Φαρμακογενωμική εξετάζει σε γενετική βάση τις ατομικές διαφοροποιήσεις στην απόκριση σε θεραπείες σε συγκεκριμένα φάρμακα. Ο στόχος της φαρμακογονιδιωματικής είναι να επιτευχθεί εξατομικευμένη θεραπεία με την πρόβλεψη της επιδεκτικότητας σε ασθένειες και την αντίδραση σε φάρμακα και εμβόλια. Τέτοια μελέτη δεν θα πρέπει να περιορίζεται σε μεμονωμένα γονίδια ή σε απλούς πολυμορφισμούς νουκλεοτιδίων (SNPs), καθώς τα γονίδια αλληλεπιδρούν μεταξύ τους. Η βιολογία συστημάτων χρειάζεται να μελετήσει τις αλληλεπιδράσεις μεταξύ των βιολογικών στοιχείων προς την κατανόηση ασθενειών σε επίπεδο συστήματος. Ως μια νέα προσέγγιση στην ανάλυση βιολογικών συστημάτων σε όλα τα επίπεδα των πληροφοριών, η συστημική βιολογία μπορεί να προσφέρει νέες στρατηγικές για την ανακάλυψη φαρμάκων και την ανάπτυξη.

Ωστόσο, ένα από τα πιο σημαντικά εμπόδια στην πρακτική της εξατομικευμένης ιατρικής είναι η μετάφραση των επιστημονικών ανακαλύψεων σε καλύτερα θεραπευτικά αποτελέσματα. Ένας κρίσιμος παράγοντας για την επιτυχή μετάφραση είναι η πρόσβαση και η ανάλυση σε ολοκληρωμένα στοιχεία εντός και μεταξύ των λειτουργικών τομέων. Για παράδειγμα, τα περισσότερα από τα δεδομένα κλινικής και βασικής έρευνας είναι σήμερα αποθηκευμένα σε διαφορετικές και ξεχωριστές βάσεις δεδομένων, και είναι αναποτελεσματικό για τους ερευνητές να έχουν πρόσβαση σε αυτά τα δεδομένα. Παρά το γεγονός ότι υπήρξε μια αυξανόμενη ζήτηση για τη διαχείριση των δεδομένων, λίγα εργαλεία είναι διαθέσιμα που ανταποκρίνονται στη ζήτηση.

Με την πληθώρα των διαφόρων τεχνολογιών, η εκθετική ανάπτυξη των ιατρικών πληροφοριών ξεπερνά κατά πολύ την ικανότητα μας να την αφομοιώσουμε και να την εφαρμόσουμε στις κλινικές. Ωστόσο, τα παραδοσιακά κλινικά πληροφοριακά συστήματα δεν έχουν εστιαστεί στη διαχείριση της γνώσης ή λειτουργίες υποστήριξης αποφάσεων. Με την ισχυρή ζήτηση για εξατομικευμένη ιατρική με βάση τη συστημική βιολογία και τη φαρμακογονιδιωματική, η ανάγκη για νέα πληροφορική υποστήριξη για τη βελτίωση της επικοινωνίας μεταξύ των βασικών επιστημόνων και νοσοκομειακών γιατρών γίνεται ολοένα και πιο επιτακτική.

Η μεταγραφική βιοπληροφορική είναι μια ισχυρή μέθοδος για τη γεφύρωση του χάσματος μεταξύ συστημικής βιολογίας και πρακτικής εξατομικευμένης ιατρικής. Βιοπληροφορική χρησιμοποιεί υπολογιστικές προσεγγίσεις για την επίλυση προβλημάτων και τη βελτίωση της επικοινωνίας, κατανόησης, καθώς και την ανάλυση και τη διαχείριση των βιοϊατρικών πληροφοριών. Όπως ορίζεται από την Αμερικανική Ένωση Ιατρικής Πληροφορικής (AMIA), η μεταφραστική βιοπληροφορική είναι έναν νέο πεδίο «βελτιστοποίησης και μετατροπής των όλο και πιο εκτενών βιοϊατρικών δεδομένων, και δεδομένων γονιδιωματικής κυρίως, σε δυναμική, έξυπνη, προληπτική και συμμετοχική υγεία .

Για παράδειγμα, τα πιο δύσκολα και κρίσιμα τμήματα σε κλινικές αξιολογήσεις περιλαμβάνουν αναποτελεσματική διαχείριση των κλινικών ροών εργασίας, αναποτελεσματική επικοινωνία, και η έλλειψη ενός κεντρικού συστήματος διαχείρισης των δεδομένων. Βιοϊατρικές εφαρμογές πληροφορικής στη μεταγραφική ιατρική θα επιτρέψουν την αποτελεσματική διαχείριση των ροών εργασίας σε αμφότερα τα κλινικά και ερευνητικά περιβάλλοντα. Η Μεταγραφική

βιοπληροφορική θα βελτιώσει την ενσωμάτωση των κλινικών και εργαστηριακών ροών δεδομένων. Εφαρμογές όπως η αρχιτεκτονική ηλεκτρονικού ιατρικού φακέλου θα διευκολύνουν τη κοινοποίηση της πληροφορίας, καθώς και δημιουργία και εφαρμογή προτύπων. Επιπλέον, η μεταγραφική βιοπληροφορική μπορεί να βοηθήσει τη μείωση των κλινικών κινδύνων και την αποτελεσματική χρήση των πόρων υγειονομικής περίθαλψης. Για παράδειγμα, οι πληροφορίες που βασίζονται σε υπολογιστικά συστήματα είναι η πιο οικονομικά αποδοτική και πολλά υποσχόμενη στρατηγική για πρόληψη των δυσμενών θεραπευτικών επιπλοκών.

Το πιο σημαντικό στοιχείο της μεταγραφικής βιοπληροφορικής είναι ότι μπορεί να προωθήσει την πρακτική της εξατομικευμένης ιατρικής. Θα εξουσιοδοτούν τους επιστήμονες και κλινικούς ιατρούς να σχεδιάσουν εξατομικευμένες στρατηγικές για να χορηγούν το σωστό φάρμακο με τις σωστές δόσεις στους σωστούς ανθρώπους (βλ.. Εικ. 1). Τέτοιες προσεγγίσεις θα βοηθήσουν να ξεπεραστεί η θεραπευτική αντίσταση και οι δυσμενείς επιπτώσεις, καθώς και βελτίωση της επικοινωνίας μεταξύ διεπιστημονικών ομάδων.

Οι στόχοι αυτοί μπορούν να επιτευχθούν μέσω της παροχής ενοποιητικών μεθόδων που επιτρέπουν μοντέλα πρόβλεψης σε θεραπευτικές αποκρίσεις. Μελέτες πληροφορικής σε συστημική βιολογία μπορούν να επιτρέψουν την προσομοίωση δικτύων αλληλεπιδρώντων συστατικών. Σε συνδυασμό με μελέτες υψηλής απόδοσης, η βιοπληροφορική μπορεί να βοηθήσει στον εντοπισμό του γενετικού προφίλ τόσο του ασθενούς όσο και υποομάδων ασθενών, προκειμένου να αναπτυχθεί η βέλτιστη θεραπεία.

Η μεταγραφική βιοπληροφορική έχει ως αποστολή να αντιμετωπίσει τις προκλήσεις στην ανάπτυξη της εξατομικευμένης ιατρικής. Οι προκλήσεις αυτές προέρχονται από δύο πλευρές, η πληροφορική πλευρά και η βιοϊατρική πλευρά, όπως απεικονίζεται στο Σχ. 1. Καθορίζουν το όραμα που οδηγεί στον τρόπο με τον οποίο η μεταγραφική βιοπληροφορική θα μελετηθεί, θα αναπτυχθεί και θα εφαρμοστεί. Από την πλευρά της βιοϊατρικής, η μεταγραφική βιοπληροφορική θα επιτρέψει τον προσδιορισμό των βιολογικών δεικτών των συστημάτων, την κατανόηση των συσχετίσεων γονότυπου-φαινότυπου, και τη μοντελοποίηση συστηματικών αλληλεπιδράσεων. Αυτά δεν μπορούν να επιτευχθούν χωρίς συγκεκριμένη βάση πληροφορικής. Ως εκ τούτου, από την πλευρά της πληροφορικής, η ενσωμάτωση, αποθήκευση και αλίευση των δεδομένων μπορεί να παρέχει υποστήριξη στη λήψη αποφάσεων τόσο για τους ερευνητές όσο και για τους κλινικούς ιατρούς.

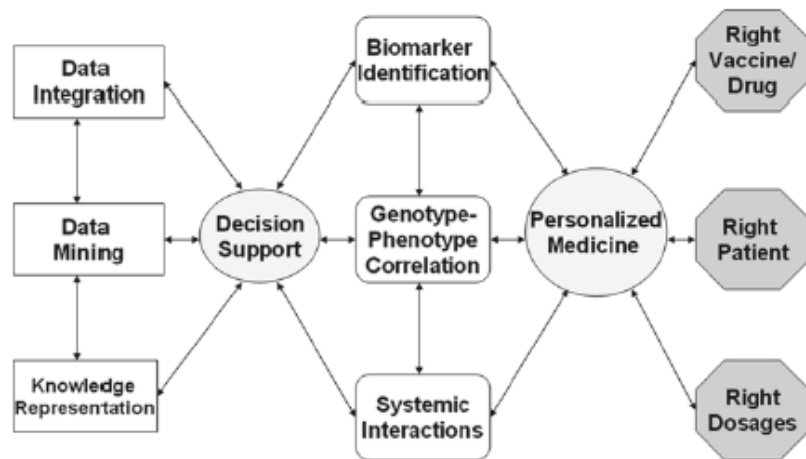


Fig. 1. The roadmap of translational bioinformatics and systems biology for the development of personalized medicine.

ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] “**Systems Biology in Drug Discovery and Development Methods and Protocols**”, Qing Yan *PharmTao, Santa Clara, CA, USA*

[2] “**Drug Design and Discovery Methods and Protocols**”, Seetharama D. Satyanarayanajois Department of Basic Pharmaceutical Sciences, College of Pharmacy, University of Louisiana at Monroe, Monroe, LA, USA

[3] Ιστοσελίδα <http://en.wikipedia.org>